

**PLASTID PROMOTERS FOR TRANSGENE EXPRESSION IN THE
PLASTIDS OF HIGHER PLANTS**

By **Pal Maliga**
Daniel Silhavy
Priya Sriraman

Pursuant to 35 U.S.C. §202(c) it is
acknowledged that the U.S. Government has certain rights
in the invention described herein, which was made in part
with funds from the National Science Foundation, Grant
5 Number MCB96-30763.

FIELD OF THE INVENTION

The present invention relates to plant genetic
engineering and particularly to plastid transformation in
10 higher plants. The invention provides novel promoter
sequences useful for the expression of foreign genes of
interest in various plant species.

BACKGROUND OF THE INVENTION

15 Chloroplast genes are transcribed by an RNA
polymerase containing plastid-encoded subunits homologous
to the α , β and β' subunits of *E. coli* RNA polymerase.
The promoters utilized by this enzyme are similar to *E.*
coli σ^{70} - promoters consisting of -35 and -10 consensus
20 elements (G.L. Igloi and H. Kossel, Crit. Rev. Plant Sci.
10, 525, 1992; W. Gruissem and J.C. Tonkyn, Crit. Rev.
Plant. Sci. 12:-19, 1993) Promoter selection by the
plastid-encoded RNA polymerase is dependent on nuclear-
encoded sigma-like factors (Link et al. 1994, *Plant*
25 *promoters and transcription factors*, Springer Verlag,
Heidelberg, pp 63-83). In addition, transcription
activity from some promoters is modulated by nuclear-
encoded transcription factors interacting with elements
upstream of the core promoter (L.A. Allison and P.
30 Maliga, *EMBO J.*, **14**:3721-3730; R. Iratni, L. Baeza, A.
Andreeva, R. Mache, S. Lerbs-Mache, *Genes Dev.* **8**, 2928,

1994, Sun et al., *Mol. Cell Biol.* **9**:5650-5659, 1989). These factors mediate nuclear control of plastid gene expression in response to developmental and environmental stimuli.

5 The existence of a second nuclear encoded polymerase transcription system in plastids has been demonstrated. However, the relevant nucleic acid sequences required for transcription initiation comprising the novel regulatory elements of this system
10 have yet to be elucidated. It is an object of the present invention to provide these novel genetic elements. Incorporation of these regulatory elements into specific plastid directed DNA constructs enables greater flexibility and range in plant species available
15 for plastid transformation, and facilitates ubiquitous expression of foreign proteins and/or RNAs and are useful in non-green plastids.

SUMMARY OF THE INVENTION

20 Promoters contain distinct DNA sequence information to facilitate recognition by the RNA polymerase and initiation of transcription leading to gene expression. In accordance with the present invention, promoters have been discovered which function
25 in both monocots and dicots. These promoter elements may be used to advantage to express foreign genes of interest in a wider range of plant species. Additionally, the promoter elements of the invention drive expression of exogenous genes in non-green tissues. It is an object of
30 the present invention to provide DNA constructs and methods for stably transforming plastids of multicellular plants containing such promoters. The DNA constructs of the invention extend the range of plant species that may be transformed.

35 The promoters recognized by plastid-encoded

plastid RNA (PEP) polymerase have been well characterized in photosynthetic tissues such as leaf. The utility of PEP promoters for expression of foreign proteins in non-green tissues is demonstrated herein. The nuclear-encoded plastid (NEP) polymerase transcription system of the present invention directs expression of plastid genes also in roots, seeds, meristematic tissue and/or leaves. In most plants, including maize, cotton and wheat, plant regeneration is accomplished through somatic embryogenesis (i.e., involving meristematic tissue). In a preferred embodiment of the invention, efficient plastid transformation in these crops will be greatly facilitated, through the use of the NEP and PEP plastid transcription system and promoters of the present invention.

Particularly preferred promoters for use in the constructs of the invention are the *clpP* -111 (SEQ ID NOS: 15, 16, 30 and 31) promoters for the transformation of monocots and dicots and the *pclp* -53 promoter for transformation in dicots. Homologous *clpP* promoters from other plant species are contemplated to be within the scope of the present invention.

Other preferred promoters for use in expressing foreign genes of interest in the plant plastid in non-green tissues are PEP promoters present in the 16SrDNA operon, SEQ ID NOS: 28 and 29. Additional promoter elements suitable for use in the present invention are the *rpoB* and *atpB* promoters.

The NEP promoters of the invention are incorporated into currently available plastid transformation vectors such as those described in pending U.S. Application No. 08/189,256, and also described by Svab & Maliga., *Proc. Natl. Acad. Sci. USA*, 90, 913 (1993). Protocols for using such vectors are described in U.S. Patent No. 5,451,513. The disclosures of the

three references cited above are all incorporated by reference herein. To obtain transgenic plants, plastids of non-photosynthetic tissues are transformed with selectable marker genes expressed from NEP promoters and transcribed by the nuclear-encoded polymerase. Likewise, to express foreign proteins of interest, expression cassettes are constructed for high level expression in non-photosynthetic tissue, using the NEP promoter transcribed by the nuclear-encoded plastid RNA polymerase. In another aspect of the invention, PEP promoters of the invention are incorporated into currently available plastid transformation vectors and protocols for use thereof.

In yet another aspect of the invention, the NEP transcription system also may be combined with the σ^{70} - type system through the use of dual NEP/PEP promoters. In transforming DNA constructs, the promoters are arrayed in tandem, operably linked to the coding region of the foreign gene of interest. As used herein, the term transcription unit refers to isolated DNA segments which comprise the essential coding regions of one or more exogenous protein(s) of interest. Such transcription units may also contain other cis elements for enhancing gene expression, such as enhancer elements.

Transcription units are operably linked to the promoters of the invention, such that expression of the transcription unit is regulated by said promoter. Particularly preferred promoters for use in combination are the Prrn PEP promoters combined with the clpP type II NEP promoter in dicots and the Prrn PEP promoter combined with the clpP Type I NEP promoter for use in both monocots and dicots. A suitable Prrn promoter has the following sequence (SEQ ID NO: 32) 5'-GCTCCCCCGC CGTCGTTCAA TGAGAATGGA TAAGAGGCTC GTGGGATTGA CGTGAGGGGG CAGGGATGGC TATATTCTG GGAGCGAACT CCGGGCGAAT ACGAAGCGct

TGGATACAGT TGTAGGGAGG GATT-3'.

Homologous PEP and NEP promoters from a variety of plant species corresponding to those listed above are also considered to fall within the scope of the present invention. The transforming DNA also contains 3' regulatory regions of plant or bacterial origin to effect efficient termination of transcription. An exemplary 3' regulatory region is shown in Figure 13, SEQ ID NO: 27.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 depicts a series of autoradiographs illustrating RNA steady state concentrations in green (G) and white (W) iojap maize leaves. To control for loading, blots shown above were reprobbed for cytoplasmic 25S ribosomal RNA as described (Dempsey et al., *Mol. Plant Path.* **83**:1021-1029, 1993) (lower panels).

Figures 2A and 2B depict a pair of autoradiographs showing the results of primer extension analysis on green (G) and white (W) iojap maize leaves for the mapping of the *clpP* promoter, Fig. 2A. The number -111 refers to the transcript 5' end relative to the ATG translation initiation codon. Figure 2B is an autoradiograph showing the results of an in vitro capping and RNase protection assay to identify primary transcript 5' ends. Note that the RNase protection construct is short and protects only a 79 nt fragment. Size of molecular weight (MW) markers (100, 200, 300, 400, and 500 nucleotides) is also shown.

Figures 3A and 3B show the mapping of maize *rpoB* promoters. Fig. 3A is an autoradiograph showing the results of primer extension analysis of RNA from green (G) and white (W) iojap maize leaves. The number -147

refers to transcript 5'-end. Fig. 3B shows the results of an in vitro capping and RNase protection assay to identify primary transcript 5'-ends. Note that the RNase protection construct is short and protects only a 74 nt fragment. Size of molecular weight (MW) markers (100, 200, 300, 400, and 500 nucleotides) is also shown.

Figures 4A, 4B and 4C show the mapping of *atpB* promoters. Fig. 4A show the results of primer extension analysis of RNA from green (G) and white (W) iojap maize leaves. Number -298 and -601 refer to transcript 5'-ends. Fig. 4B shows the results of an in vitro capping and RNase protection assay to identify primary transcript 5' ends. Note that the RNase protection construct is short and protects only a 79 nt fragment. Size of molecular weight (MW) markers (100, 200, 300, 400, and 500 nucleotides) is also shown. Fig. 4C depicts a physical map of the *atpB* - *rbcL* intergenic region. Map position of the primary transcript 5' ends for NEP and PEP promoters are marked with filled and open circles, respectively.

Figure 5A shows the alignment of DNA sequences flanking the NEP promoter transcription initiation sites. Site of transcription initiation is marked (filled circles). Regions with significant similarity are boxed (Box I and Box II). Sequences corresponding to the loose 10-nt dicot NEP consensus are underlined with thin lines. *clpP* -111 promoter region in maize is indicated by the thick underlining. Figure 5A shows *rpoB* (SEQ ID NO: 1), *atpB* (SEQ ID NO: 2) and *clpP* (SEQ ID NO: 3) promoter regions, respectively in maize. The alignment and nucleotide sequences of *atpB* promoters in maize, sorghum, barley, wheat, and rice are shown in Figure 5B and correspond to SEQ ID NOS: 4, 5, 6, 7 and 8 respectively. Figure 5C shows the alignment of the *rpoB* sequences in

maize (SEQ ID NO: 9), rice (SEQ ID NO: 10) and tobacco (SEQ ID NO: 11). The alignment of the clpP sequences in maize (SEQ ID NO: 12), rice (SEQ ID NO: 13), and wheat (SEQ ID NO: 14) is shown in Figure 5D.

5

Figure 6 shows the sequence alignment of the tobacco (SEQ ID NO: 15) and rice (SEQ ID NO: 16) plastid clpP promoter regions. The NEP transcription initiation sites are marked with filled circles, the PEP transcription initiation site is marked with an open circle. The third tobacco NEP promoter is outside the sequence shown. The clpP coding regions are boxed. The 29-bp shared homologous region around the Type II tobacco PclpP-53 promoter is underlined.

10

15

Figure 7 is a schematic drawing of the plastid targeting region of plasmid pDS44. On top is shown the restriction map of the chimeric uidA gene. PclpP is the rice clpP promoter fragment between the SacI and Nco site (SEQ ID NO: 17) which is shown at the bottom of Fig. 5. uidA encodes the β -glucuronidase reporter enzyme. Trps16 is the rps16 ribosomal protein gene 3' untranslated region. A map of the transforming DNA in a pPRV111A plastid vector (Gene Bank Accession No. U12812) is also shown (Zoubenko et al., Nucleic Acids Res. 22:3819-3824, 1994).

20

25

Figures 8A and 8B illustrate the promoter activity of the rice clpP promoter region in transgenic tobacco plastids. Figure 8A shows the results of primer extension analysis to map RNA 5'-ends upstream of the rice PclpP::uidA::Trps16 chimeric gene. Primer extension analysis was carried out on total cellular RNAs isolated from the leaves of wild-type (wt) and transplastomic (T) tobacco plants. The numbers (-61, -111, -136, -169, 177)

30

35

refer to nucleotide positions of the mapped 5'ends
relative to the ATG translation initiation codon of the
rice clpP gene. Figure 8B shows a schematic
representation of RNA 5' ends mapped in the rice clpP
promoter region in rice (Os) and in tobacco (Os in Nt),
and upstream of the wild-type tobacco clpP gene (Nt).
Promoters identified in homologous systems are marked
(NEP, filled circles, PEP, open circles). Numbers
indicate distance from translation initiation codon
(nucleotide upstream of ATG is position -1).

Figure 9 depicts an alignment of DNA sequences
which are conserved around the plastid clpP transcription
initiation sites. Sequences are aligned for *Marchantia*
polymorpha (Kochi) (SEQ ID NO: 18), *Pinus contorta*
(Clarke) (SEQ ID NO: 19), spinach (Westhof) (SEQ ID NO:
20), tobacco (Hajdukiewicz et al., 1977) (SEQ ID NO: 21),
rice (SEQ ID NO: 22), maize (SEQ ID NO: 23) and
Arabidopsis (SEQ ID NO: 24). Wild-type RNA 5' ends are
marked with filled circles. The -61 5' end of the rice
clpP-uidA chimeric mRNA is marked with an asterisk.
Conserved nucleotides are boxed if identical at least in
five species. The translation start codons are shaded.

Figure 10 shows the plastid targeting region of
plasmid pPS18. Plasmid pPS18 has the plastid targeting
region of plastid vector pPRV111A, GenBank Accession No:
U12812, with a chimeric uidA gene expressed from the
PclpP-53(-22/+25) promoter. DNA sequence of the uidA gene
is shown in Figure 13 (SEQ ID NO: 27).

Figure 11 depicts the clpP 5'fragments tested
for promoter activity. The largest segment contains the
PclpP-53 and PclpP-95 transcription initiation sites
derived from Type II NEP and PEP promoters, respectively.

The boundaries give the distance in nucleotides from PclpP-53 transcription initiation site (+1). SEQ ID NO: 25 contains the sequences present between -22 and +25 and SEQ ID NO: 26 shows the nucleotides present between -10 and +25. Both sequences function as promoters.

Figure 12 depicts the results of primer extension analysis to test promoter activity of clpP 5' fragments included in Figure 11. Transcripts derived from PclpP-53 (-53), a minor NEP promoter (*), PclpP-95 (-95) and PclpP-173 (-173) are marked. DNA sequences on the side give distance from primer in nucleotides.

Figure 13 shows the DNA sequence (SEQ ID NO: 27) of the chimeric uidA gene. SacI and HindIII cloning sites are marked. SacI, XhoI, NcoI, XbaI and HindIII sites are underlined. Translation initiation (ATG) and stop (TGA) codons are underlined twice. Nucleotides derived from the tobacco plastid genome are in capital letters; the position of the first and last nucleotide within the genome is listed.

Figure 14 is a northern blot showing RNA steady state concentrations in leaf (L) and in embryogenic cultured cells (E) of rice. To control for loading, the blots were stripped and probed for cytoplasmic 25S ribosomal RNA (lower panels). Hybridization signals were quantified with a Molecular Dynamics PhosphorImager. The fold excesses in leaves over cultured cell signals are shown below the panels.

Figure 15 is a Southern blot showing the relative plastid genome copy number in leaves (L) and in cultured embryogenic cells (E) of rice. EcoRI-digested total cellular DNA (approximately 2 µg per lane) was

probed for plastid 16SrDNA (upper) and nuclear-encoded 25SrDNA (lower). Hybridization signals were quantified with a Molecular Dynamics PhosphorImager. The ratio of 16SrDNA/25SrDNA signal intensity in leaves relative to embryogenic cells was 1.1.

Figure 16 is an autoradiograph depicting the mapping results of plastid mRNA promoters in rice leaves (L) and embryogenic cultured cells (E) using primer extension analysis. Numbers on the right indicate the distance between the translation initiation codon (ATG) and 5' ends of primary transcripts (PEP, ○; NEP, ●), or of processed mRNAs (-). DNA sequences on the left are size markers. Sequence ladders shown for 16S rDNA and clpP were obtained with homologous template and oligonucleotides used for primer extension analysis.

Figure 17 shows the alignment of the maize (SEQ ID NO: 28 and rice (SEQ ID NO: 29) PEP promoter in the rrn operon, Figure 17A. Figure 17B shows the alignment of the plastid clpP promoter regions in maize (SEQ ID NO: 30) with the homologous regions in rice (SEQ ID NO: 31). PEP (○) and NEP (●) transcription initiation sites and processed 5' ends (-) are marked. For 16SrDNA sequence information see Strittmatter et al. (1985). The present invention provides sequence information for maize clpP.

DESCRIPTION OF THE INVENTION

Several reports have suggested the existence of an additional plastid-localized, nuclear-encoded RNA polymerase (reviewed in Grussem and Tonkyn, 1993; Igloi and Kossel, 1992; Mullet, 1993; Link, 1994). By deleting the rpoB gene encoding the essential β subunit of the tobacco E. coli-like RNA polymerase, the existence of a second plastid transcription system which

is encoded by the nucleus has been established (Allison et al., 1996, EMBO J. **15**:2802-2809). Deletion of *rpoB* yielded photosynthetically defective, pigment-deficient plants.

5 While the activity of the previously-known plastid-encoded σ^{70} -type transcription system in photosynthetically active tissues, such as leaf, has been the subject of much research, the nuclear-encoded polymerase transcription system has not yet been
10 characterized. In accordance with the present invention, it has been discovered that the NEP system also directs expression in roots, seeds and meristematic tissue. In most plants, including maize, cotton and wheat, plant regeneration is accomplished through somatic
15 embryogenesis (i.e., involving meristematic tissue). Efficient plastid transformation in these crops will be enabled, or greatly facilitated, through the use of *clpP* promoters driven by the NEP plastid transcription system of the present invention.

20 The ATP dependent Clp protease is widespread, if not ubiquitous, among both procaryotic and eucaryotic cells (Goldberg, Eur. J. Biochem. 203:9-23, 1992). Clp protease expression has recently been examined in *chlamydomonas* (Huang et al. Mol. Gen. Genetics 244:151-
25 159, 1994). The results of gene knockout experiments in this unicellular algae demonstrated the following: 1) all the transformants were found to be heteroplasmic mutants containing both the disrupted *clpP* and wild-type copies; 2) the transformants persisted as heteroplasmic
30 mutants after six rounds of growth and screening under selection pressure for the disrupted *clpP*; and 3) the heteroplasmic mutant stabilized at a level where approximately 80% of the *clpP* DNA copies were disrupted. These data indicate that the *clpP* is essential for cell
35 growth even under conditions where photosynthesis is not

required.

Shanklin et al. have examined the immunolocalization of clpP protein in Arabidopsis chloroplasts (Shanklin et al. The Plant Cell 7:1713-1722, 1995). These studies revealed that clpP and clpC are constitutively expressed in all tissues of Arabidopsis at levels equivalent to those of E. coli clpP and clpA. The observation that the clpP NEP promoter drives constitutive gene expression in all parts of a plant makes this promoter particularly suitable for use in the present invention.

The NEP promoters of the invention are incorporated into currently available plastid transformation vectors such as those described in pending U.S. Application No. 08/189,256, and also described by Svab & Maliga., *Proc. Natl. Acad. Sci. USA*, 90, 913 (1993). Protocols for using such vectors are described in U.S. Patent No. 5,451,513. The disclosures of the three references cited above are all incorporated by reference herein. To obtain transgenic plants, plastids of non-photosynthetic tissues as well as photosynthetic tissues are transformed with selectable marker genes expressed from NEP promoters and transcribed by the nuclear-encoded polymerase. Likewise, to produce foreign proteins of interest, expression cassettes are constructed for high level expression in non-photosynthetic tissue, using the clpP NEP promoter transcribed by the nuclear-encoded plastid RNA polymerase. The NEP transcription system also may be combined with the σ^{70} -type system through the use of dual NEP/PEP promoters.

For versatility and universal applications, expression of selectable marker genes for plastid transformation is desirable in all targeted tissue types at a high level. Selectable marker genes in the

currently utilized plastid transformation vectors are expressed from PEP promoters recognized by the plastid encoded RNA polymerase. The PEP polymerase transcribes photosynthetic genes and some of the housekeeping genes, and therefore appears to be the dominant RNA polymerase in photosynthetically active leaf tissues. Efficient plastid transformation has been achieved in tobacco based on chloroplast transformation in leaf cells. However, plant regeneration is not feasible, or is not practical from the leaves of most agronomically important cereal crops, including maize, rice, wheat and in cotton. In these crops, transgenic plants are typically obtained by transforming embryogenic tissue culture cells or seedling tissue. Given that these tissues are non-photosynthetic, expression of marker genes by constitutive clpP NEP promoters which are active in non-green tissues will facilitate transformation of plastids in all non-photosynthetic tissue types. Furthermore, as demonstrated herein, the ribosomal RNA operon PEP promoter is highly active in rice embryogenic cells.

The following nonlimiting Examples describe the invention in greater detail. Specifically, Examples I-III below describe preferred methods for making and using the DNA constructs of the present invention and for practicing the methods of the invention. Any molecular cloning and recombinant DNA techniques not specifically described are carried out by standard methods, as generally set forth, for example, in Ausubel (ed.), Current Protocols in Molecular Biology. John Wiley & Sons, Inc. (1994).

EXAMPLE I

Identification of promoters for the nuclear-encoded plastid RNA polymerase in the

ribosome-deficient maize mutant iojap

As described previously, plastid promoters with conserved -10/-35 elements are well characterized for the plastid-encoded plastid RNA polymerase, PEP.

Additionally, a ten-nucleotide promoter consensus was reported for a second, nuclear encoded, plastid RNA polymerase, NEP, in tobacco, a dicot (Hajdukeiwicz et al. 1997, in press). In this Example, NEP promoters active in monocots are described. NEP promoter mapping was carried out in the plastid ribosome-less maize mutant iojap which lacks PEP. These studies have revealed that atpB, an ATPase subunit gene, contains promoters for both NEP and PEP. In contrast, clpP, a protease subunit gene, and the rpoB operon, encoding the rpoB, rpoC1 and rpoC2 PEP subunit genes, are transcribed from NEP promoters only. These findings suggest conservation of transcription systems between monocots and dicots for the expression of certain plastid genes. The monocot NEP promoters share sequence homology around the transcription initiation site, including the 10-nucleotide loose consensus identified in dicots.

In the plastids of photosynthetic higher plants, genes are transcribed by at least two RNA polymerases: the plastid-encoded plastid RNA polymerase (PEP) and the nuclear-encoded plastid RNA polymerase (NEP). The sigma-factor homologues and PEP regulatory factors are encoded in the nucleus and are imported into plastids (Igoli and Kossel, (1992) supra; Link, (1996) Bioessays 18:465-471; Stern DB, Higgs DC, Yang J (1997) Trends Plant Sci 2:308-315). Much less is known about NEP. One appealing candidate for NEP is a 110 kD protein which has biochemical properties similar to yeast mitochondrial and T7 RNA polymerases. NEP may require

accessory factors, such as CDF2 for its activity
(Lerbs-Mache, 1993; Iratni et al., 94; Genes and dev).

Dicot promoters for NEP were identified in
tobacco plants lacking PEP due to deletion of rpoB
5 encoding the PEP β subunit. The general rule emerging
from these studies is that plastid genes fall into three
classes. Class I genes contain only PEP promoters.
Examples of this class are the photosytem I and II genes.
Class II genes contain both NEP and PEP promoters.
10 Representative members of this class of genes are genes
involved in plant metabolism and housekeeping genes.
Genes in the third class contain NEP promoters only.
Genes in this class include rpoB and accD.

As plastid transformation is not yet available
15 in monocots, Δ rpoB plants could not be obtained for maize
NEP promoter analysis. However, mutants with a defect in
plastid ribosome accumulation are available in barely
(albostrians; Hess et al., *EMBO J.* **12**:563-571, 1993) and
maize (iojap; Walbot and Coe, *Proc. Natl. Acad. Sci.*
20 **76**:2760-2764, 1979; Han et al., *Planta* **191**:552-563, 1993;
Han et al., *EMBO J.* **11**:4037-4046, 1992). In the absence
of plastid ribosomes, these mutants cannot synthesize
PEP. Both mutants accumulate mRNAs for a subset of
plastid genes indicating the presence of NEP activity
25 (Hess et al., 1993, *supra*; Han et al., 1993, *supra*).

NEP promoters for three plastid genes were
mapped in white ribosome-less maize iojap seedlings.
The data show that the maize atpB gene has alternative
NEP and PEP promoters while the clpP and the rpoB genes
30 are transcribed from NEP promoters exclusively in both
white and green seedlings. DNA sequence alignment
revealed that monocot NEP promoters share homology
directly upstream of the transcription initiation site.
The homologous region includes the previously identified
35 tobacco NEP promoter consensus elements suggesting

conservation of the NEP transcription machinery between monocots and dicots.

Materials and Methods for Example I

Plant Materials. Ioja is a recessive striped mutant of maize. Maternal white and green seedlings were obtained by crossing a striped *ij/ij* maternal parent (1404) with pollen from a wild type male (inbred Oh51a). The seeds were kindly provided by Rob Martienssen and Mary Byrne, Cold Spring Harbor Laboratory. Surface-sterilized seeds were germinated in vitro on 2% MS medium (24 °C, 16 hours illumination).

RNA Gel blots. Total cellular RNA was prepared from the leaves of 9-day-old seedlings (Stikema et al. *Plant Mol. Biol.* **11**:255-269, 1988). The RNA (5 µg per lane) was electrophoresed on 1% agarose/formaldehyde gels, then transferred to Hybond N (Amersham) using the Posiblote Transfer apparatus (Stratagene). Hybridization to random-primer labeled fragment was carried out in Rapid Hybridization Buffer (Amersham) overnight at 65 °C.

Double-stranded ptDNA probes were prepared by random-primed ³²P-labeling of PCR-generated or gel-purified DNA fragments. The sequence of the primers used for PCR, along with their positions within the tobacco (**N.t.; Genbank Accession No. Z00044**) or maize (**Z.m.; Genbank Accession No. X86563**) ptDNA are as follows:

Gene	5' nt position in ptDNA	Sequence	SEQ ID NO:
atpB (Z.m.)	55860 (C)	GAGAGGAATGGAAGTGATTGACA	(33)
	55103	GAGCAGGGTCGGTCAAATC	(34)
clpP (Z.m.)	69840	ATCCTAGCGTGAGGGAATGCTA	(35)
	70064 (C)	AGGTCTGATGGTATATCTCAGTAT	(36)
psbA (N.t.)	1550 (C)	CGCTTCTGTAAGTGG	(37)

The following ptDNA fragments were used as probes: rbcL(N.t.), a BamHI fragment (nucleotides 58047 to 59285 in ptDNA); 16SrDNA (N.t.), EcoRI to EcoRV fragment (nucleotides 138447 to 140855 in the ptDNA); rpoB, HindIII fragment (nt 24291-24816).

Primer extension analysis. Primer extension reactions were carried out on 10 µg of total leaf RNA as described (Allison et al. *EMBO J.* **15**:2802-2809, 1996). The primers are listed below, with nucleotide positions in the published maize plastid genome sequence (Maier et al., *J. Mol. Biol.* **251**:614-628, 1995). Underlined oligonucleotides (added to create cloning site) were also used to generate the capping constructs, in which case the position of the first nucleotide in the genome sequence is positioned immediately following the underline.

Gene	5' nt position in ptDNA	Sequence	SEQ ID NO:
clpP#1	70182	GGTACTTTTGGAACACCAATGGGCAT	(39)
atpB#1	56095	GAAGTAGTAGGATTGGTTCTCATAAT	(40)
atpB#2	56640	<u>GGTCTAGA</u> ATTCCTATCGAATTCCTTC	(41)
rpoB#1	21545(C)	GAATCTACAAAATCCCTCGAATTG	(42)
rpoB#2	21418(C)	ACTCTTCATCAATCCCTACG	(43)

(Note: C at 3'-end of the atpB#2 oligonucleotide is at nt position 56644; the published sequence has a 15 nucleotide deletion relative to the sequence we found).

Identification of primary transcripts by in vitro capping. Total leaf RNA (20 µg for clpP or 100 µg for rpoB and atpB transcripts) from white seedlings was capped in the presence of 0.25 or 1.0 mCi [α -³²P] GTP (Kennell and Pring, *Mol. Gen. Genet.* **216**:16-24, 1989). Labeled RNAs were detected by ribonuclease protection

(Vera et al., *Plant Mol. Biol.* **19**:309-311, 1992) using the RPAII kit (Ambion). To prepare the protecting complementary RNA, an appropriate segment of the plastid genome was PCR-amplified using the primers listed below.

Gene	5' nt position	Sequence	SEQ ID NO:
	<u>in ptDNA</u>		
clpP#2	70241	<u>GGTCTAG</u> ACTTAACTTTAATATGGA	(44)
clpP#3	70549 (C)	<u>GGGAATTCTGTTT</u> GTAAGAAGA	(45)
atpB#2	56640	<u>GGTCTAGA</u> ATTCCTATCGAATTCCTTC	(46)
atpB#3	56832 (C)	<u>GGCTCGAGGGACA</u> ACTCGATAGGATTAGG	(47)
rpoB#3	21394 (C)	<u>GGTCTAGA</u> ATCTAGCAATCATGGAATC	(48)
rpoB#4	21066	<u>GGCTCGAGCGTGC</u> TATTCTAAATCGT	(49)

The 5' primers set forth above were designed to add a Xba I restriction site (underlined) upstream of the amplified fragment. The 3' primers were designed to add a XhoI (atpB , rpoB) or EcoRI (clpP) site (underlined) downstream of the amplified sequence. The amplified product was cloned after digestion with the appropriate restriction enzyme into a pBSKS+ vector (Stratagene). To generate unlabeled RNA complementary to the 5' end of RNAs, the resulting plasmid was linearized with XhoI (atpB, rpoB) or EcoRI (clpP) and transcribed in a Megascript (Ambion) reaction with T7 RNA polymerase. Markers (100, 200, 300, 400, and 500 nucleotides) were prepared with the RNA Century Markers Template Set (Ambion), following the manufacturer's protocol.

DNA sequence analysis. DNA sequence analysis was carried out utilizing the Wisconsin Sequence Analysis Package (Genetics Computer Group, Inc.).

Results and Discussion

Plastid transcript accumulation in the maternal white and green maize plants.

Lack of 16S rRNA accumulation in the white

maize plants confirmed the reported lack of plastid ribosomes in the white iojap seedlings (Walbot and Coe 1979, *supra*). The absence of *rbcL* and *psbA* mRNAs, known to be transcribed from PEP promoters, indicates that the maternal white plants indeed lack PEP activity. (Han et al., 1993, *supra*). See Figure 1.

Transcript accumulation for three additional plastid genes: *clpP*, *rpoB* and *atpB* has been analyzed in iojap plants. High steady-state mRNA levels, indicating the presence of NEP promoters was reported for these genes in iojap maize (Han et al., 1993, *supra*), albobstrians barley (Hess et al., 1993, *supra*) and $\Delta rpoB$ tobacco (Hajdukiewicz et al., *EMBO J.*, in press, 1997), respectively. Readily-detectable accumulation of mRNA in the ribosome-less iojap plants confirmed that active NEP promoters regulate expression for each of these genes (Fig. 1).

The *clpP* NEP promoter is efficiently transcribed in white and green seedlings.

To identify the NEP promoters, transcript 5'-ends were mapped by primer extension analysis. To distinguish between 5'-ends that represent transcripts from NEP promoters from those generated by RNA processing, the 5'-ends were capped using guanylyltransferase.

For *clpP*, significant mRNA accumulation was found in both white and green seedlings. Primer extension analysis identified only one major *clpP* 5'-end at nucleotide position -111 (the nucleotide upstream of the ATG being at -1 position). The *clpP* -111 transcript could be capped in vitro. See Figure 2B. These data confirm that this 5' end is a primary transcript and also identifies the maize NEP promoter *PclpP*-111. The same 5'-end is observed in both white and green maize seedlings indicating that the same *clpP* promoter is active in chloroplasts as well as in the

nonphotosynthetic iojap plastids (Fig. 2A). Therefore, PclpP -111 is considered to be a constitutive promoter. In rice, the clpP transcript 5'-end mapped to the same nucleotide indicating conservation of PclpP -111 in monocots (data not shown).

The rpoB NEP promoter activity is enhanced in iojap plastids.

RNA gel blot analysis shows that rpoB mRNA accumulates to a detectable level only in white seedlings (Fig. 1). However, more sensitive primer extension analysis revealed that the same 5'-ends were present in both white and green leaves. Two 5'-ends could be identified, a major band at nucleotide position -147, and a minor band at position -285 (Fig. 3A). The in vitro capping assay confirmed that the -147 RNA species is a primary transcript (Fig. 3B), and therefore the product of PrpoB-147 promoter.

The atpB gene is transcribed from a NEP promoter in white plants and from a PEP promoter in green seedlings.

There is substantial atpB mRNA accumulation in green leaves, while much less is found in the white iojap leaves (Fig. 1). Primer extension analysis of mRNA from green leaves identified a transcript with a 5'-end at nucleotide position -298 (Fig. 4A) confirming an earlier report (Mullet et al., *Plant Mol. Biol.* **4**:39-54, 1985). This -298 mRNA species was absent in leaf RNA isolated from white plants, indicating that the -298 mRNA is a PEP transcript. Instead, another atpB transcript was mapped to nucleotide position -601 (Fig. 4A). The difference in the size of the two mRNAs is apparent on the RNA gel blot shown in Fig. 1.

The -601 transcript could be capped in vitro by guanylyltransferase, indicating that it is a primary transcript (Fig. 4B). Therefore, it is the product of the

PatpB-601 NEP promoter, with readily detectable activity only in white iojap leaves.

5

Sequence conservation around the monocot NEP transcription initiation sites.

10 Alignment of the maize clpP, rpoB and atpB NEP promoters revealed significant homology upstream of the transcription initiation sites. In a 13-nucleotide region 8 nucleotides are shared in all three promoters (Fig. 5; Box I). In a pairwise comparison, atpB/clpP, atpB/rpoB and clpP/rpoB share 11, 10 and 8 nucleotides,
15 respectively. In addition, atpB and clpP share 8 out of 9 nucleotides further upstream (-21 to -30 relative to transcription initiation site; Box II in Fig. 5).

20 Interestingly, each maize NEP promoter has sequence similarity around the transcription initiation site with the loose dicot NEP promoter consensus CATAGAATA/GAA (Hajdukiewicz et al., 1997, *supra*; underlined in Fig. 5). For clpP, 9 nucleotides are conserved out of 10; for atpB and rpoB, the number of conserved nucleotides is 7 out of 10 (Fig. 5). In
25 addition, a second conserved region (Box II, Figure 5) is found upstream of Box I in atpB and clpP, but not in the rpoB promoters. Interestingly, the monocot Box II contains truncated versions of the dicot NEP consensus in a direct orientation: 7 out of 10 bp match in case of the
30 maize clpP (ATAGAAT) and atpB (AT-GAATA) genes (Fig. 5). These tandem repeats may play a role in regulating NEP promoter activity.

35 The regions containing the tentative maize NEP promoter sequences (-30 to +25) have been aligned with homologous regions from other monocot plants. See Figure 5B-5D. The high degree of sequence conservation indicates the presence of functional promoters in each of

these monocots species. Promoter activity for the rice homologue of the maize PclpP-111 promoter has been confirmed by primer extension analysis (data not shown).

5 The data reported herein show that maize plastid NEP promoter regions share sequence homology around the transcription initiation site with the conserved CATAGAATA/GAA NEP sequence motif in tobacco (Fig. 5). Therefore, these promoters are considered to be Type I NEP promoters. This finding indicates
10 conservation of the NEP transcription machinery between monocots and dicots. Sequences upstream of the transcription initiation sites are conserved more extensively than downstream sequences among the maize clpP, rpoB, and atpB promoters, as shown in Figure 5, similar to the dicot Type I NEP promoters (Hajdukiewicz
15 et al., 1997, *supra*).

Both, the maize PclpP-111 promoter and the tobacco PclpP-53 promoter are constitutive. In contrast to the maize promoter, the clpP promoter region in
20 tobacco lacks the CATAGAATA/GAA sequence motif (Hajdukiewicz et al., 1997, *supra*) suggesting recognition by a different NEP specificity factor (Type II NEP promoter). Interestingly, in Type II NEP promoters sequences downstream of the transcription initiation
25 sites are conserved more extensively than upstream as described in the following example. The tobacco PclpP-53 promoter homologues are the only known examples for plastid Type II NEP promoters. They have been highly conserved during evolution, including the liverworth
30 *Machantia polymorpha* and the conifer *Pinus contorta*. Although DNA sequences required for clpP Type II NEP promoter function are maintained, this region is transcriptionally silent in maize, rice and wheat. Lack of transcription from this region in cereals is probably
35 due to the loss of Type II recognition specificity. See Example II. Accordingly, the tobacco (dicot) Type II clpP promoter is suitable to drive the expression of

plastid transgenes only in dicots, whereas the cereal Type I promoter may be useful in both monocots and dicots.

The *rpoB* operon is one of few genes known to be expressed by NEP only. The *PrpoB*-147 is a Type I NEP promoter but, unlike *clpP* and *atpB*, lacks Box II (Fig. 5). Accumulation of mRNA from this promoter is low in mature leaves due to down-regulation of transcription rates (Baumgartner et al., *Plant Physiology* **101**:781-791, 1993). The *PrpoB* -147 promoter identified in this study probably plays a central role in plastid development since it regulates expression of four out of the five plastid-encoded PEP subunits (Shimada et al., *Mol. Gen. Genet.* **221**:395-402, 1990). According to one model, the two RNA polymerases form a developmental cascade during chloroplast differentiation. During the early stages of plastid development, plastid genes encoding the plastid's transcription and translation apparatus would be transcribed NEP. Once PEP is made, it would initiate transcription of photosynthetic genes from PEP promoters, and take over transcription of housekeeping genes from alternative PEP promoters (Hess et al., 1993, *supra*; Lerbs-Mache, *Proc. Natl. Acad. Sci.* **90**:5509-5513, 1993; Mullet, *Plant Physiol.* **103**:309-313, 1993; Hajdukiewicz et al. 1997, *supra*). Consistent with this model is transcription of *rpoB* from a NEP promoter. However, in maize at least one gene, *clpP*, is exclusively and efficiently transcribed by NEP in mature chloroplasts indicating that NEP remains active and essential for cellular functions even if PEP is present. An alternative hypothesis is proposed herein, which assumes that NEP and PEP are constitutively present all the time and selective transcription is mediated by promoter-specific transcription factors. The identification of NEP promoters for dicots (Hajdukiewicz et al., 1997, *supra*) and monocots (described herein) facilitates the elucidation of the roles these two plastid RNA

polymerases play in plastid function and development.

EXAMPLE II

Altered clpP Promoter Recognition by the Nucleus-Encoded Plastid RNA Polymerase Suggests Loss of a Conserved Plastid Transcription Factor in Monocots

The plastid clpP gene is transcribed by the nuclear-encoded plastid RNA polymerase (NEP) in rice, a monocot, and tobacco, a dicot. However, the two NEP promoters do not share sequence homology. To assess conservation of NEP promoter recognition between monocots and dicots, a reporter gene (uidA) expressed from the rice clpP promoter region has been introduced into tobacco plastids. The data indicate that in tobacco, transcription initiates at the correct site from the rice clpP promoter. Thus, NEP promoter recognition for this gene is conserved in both monocots and dicots. Surprisingly, transcription from the rice sequence initiated at a second site, which possesses a short stretch of homologous sequence similar to the tobacco clpP promoter region. Sequences around the clpP transcription initiation site are conserved in tobacco, a dicot, *Marchantia polymorpha*, a bryophyte, and *Pinus contorta*, a conifer. Lack of transcription from this region in rice and other cereals indicates the evolutionary loss of a factor required for NEP Class II promoter specificity.

Materials and Methods for Example II

Construction of plasmid pDS44.

Plasmid pDS44 is a pLAA24 derivative (Zoubenko et al., *Nucleic Acids Res.* **22**:3819-3824, 1994) which carries a uidA reporter gene expressed from a Pr_{rrn} promoter. Plasmid pDS44 was obtained by excising the Pr_{rrn} promoter as an SacI/EcoRI fragment and replacing it with the rice clpP promoter region engineered as a

SacI/EcoRI fragment. The 251 nucleotide SacI/EcoRI DNA fragment containing the rice clpP promoter region (including 19 basepairs of the coding region) was obtained by PCR amplification. The sequence of the PCR primers, and the position of their first nucleotide (or of its complement) in the rice plastid genome (Hiratsuka et al., *Mol. Gen. Genet.* **217**:185-194, 1989; GeneBank Accession No. X15901) are:

P1 68520(C) gggagcTCGAATCACCATTCTTT SEQ ID NO: 50

P2 68270 gggaattcTTGGAACACCAATGGGCAT SEQ ID NO: 51

Nucleotides derived from the plastid genome are in capital letters; those included to create a restriction site are in lower case letters. SacI or EcoRI restriction sites are underlined.

Tobacco plastid transformation. Plastid transformation and regeneration of transgenic tobacco plants was carried out according to the protocol described by Svab and Maliga (*Proc. Natl. Acad. Sci.* **90**:913-917, 1993). Briefly, tobacco leaves were bombarded with tungsten particles coated with plasmid pDS44 DNA using the DuPont PDS 1000He Biolistic gun. Transgenic shoots were selected on RMOP medium containing 500 µg/ml spectinomycin dihydrochloride. Putative primary transformants were identified by histochemical staining for β-glucuronidase activity encoded by uidA (Jefferson, In *Genetic Engineering*, Vol. 10, Settlow, J.K., ed. Plenum Press, NY & London, pp 247-263, 1988). A uniform population of transformed plastid genomes was verified by Southern analysis (data not shown).

Primer extension analysis to map RNA 5'-ends.

Total leaf RNA was isolated from leaves of in vitro grown plants by the method of Stiekema et al., *supra*. Primer extension reactions were carried out on 20 µg of RNA with primer uidA PE1 as described by Allison and Maliga (1995) using primer P3: 5'-GGCCGTCGAGTTTT

TTGATTTTCACGGGTTGGGG-3' (SEQ ID NO:52) (which is complementary to the uidA coding region.

5 *DNA Sequence Analysis.* DNA sequence analysis was carried out utilizing the Wisconsin Sequence Analysis Package (Genetics Computer Group, Inc.) as described in Example IV.

10 Results and Discussion

Construction of Transgenic Plants. The sequence of rice clpP upstream region included as a promoter fragment in plasmid pDS44 is shown in Figure 6. In rice, this region contains the Os-PclpP-111 Type I NEP
15 promoter. The cognate sequence in tobacco contains two NEP promoters and one PEP promoter. The Os-PclpP-111 NEP promoter was cloned upstream of a uidA coding region (encoding β -glucuronidase or GUS) with a ribosome binding site, and the rps16 3'-untranslated region (Trps16) for
20 stabilization of the mRNA shown in Figure 7. The chimeric uidA gene was cloned next to a selectable spectinomycin resistance (aadA) gene in the pPRV111A plastid vector, Genbank Accession No. U12812, and introduced into tobacco plastids. A schematic drawing of
25 the vector is shown in Figure 7. Plastid transformants were selected on spectinomycin medium. Out of 25 spectinomycin resistant lines, 12 were GUS positive. Integration of uidA at the target site was confirmed by Southern analysis (data not shown).

30 **Primer extension analysis to test promoter activity.** Primer extension analysis was carried out to map uidA 5'-ends initiating in the Os-PclpP-111 promoter region. Three major and two minor uidA transcripts were identified in the leaves of transgenic plants (Figure
35 8A). A major transcript 5'end mapped to nucleotide position -111, the same position as in rice. This result indicates that the rice Type I PclpP-111 promoter is

properly recognized in tobacco, a dicot, indicating the broad applicability of this promoter in a variety of plant species. The second major transcript with the most intense signal was found at the rice -61 position. This transcript 5'-end mapped to a rice sequence with a 29 nucleotide stretch of homology to the tobacco Nt-PclpP-53 promoter region (Figure 6). This was unexpected, since this region in rice and maize is transcriptionally silent (Example I). The third major transcript mapped to position -136 which falls within the upstream monocot NEP box (Example I).

The two minor transcript 5'-ends mapped to nucleotide positions -169 and -177 (Figure 8). These transcripts are absent in rice leaf RNA, and they do not correspond to any of the known tobacco clpP transcripts. Therefore, these 5'-ends may be primary transcripts of fortuitous PEP or NEP promoters, or are processed mRNA 5'-ends.

Sequence alignment of the regions containing the clpP promoters The rice clpP promoter fragment was transcriptionally active in tobacco. This promoter sequence corresponds to the tobacco Type II PclpP-53 promoter. In this region, rice and tobacco share a 29 nucleotide stretch of homologous sequence with 23 conserved nucleotides. To test the conservation of this same region during evolution, sequences around the clpP transcription initiation sites were aligned, including those of the liverwort *Marchantia polymorpha* (Kohchi et al., *Curr. Genet.* **14**:147-154, 1988), the conifer *Pinus contorta*, (Clarke et al., *Plant Mol. Biol.* **26**:851-862, 1994), dicots tobacco (Hajdukiewicz et al., 1997, *supra*) spinach, *Arabidopsis* (Westhoff, *Mol. Gen. Genet.* **201**:115-123, 1985) and the monocots maize and rice. See Figure 9. With the exception of tobacco, a single transcription initiation site was mapped for each of the clpP genes. We have found, that the 29-nucleotide segment around the

clpP transcription initiation site is conserved,
(underlining) suggesting that clpP is transcribed by NEP
in all of these species.

The data presented in this Example show that
the rice Os-PclpP-111 NEP promoter is properly recognized
in tobacco plastids. Transcription from the rice
PclpP-111 NEP promoter in tobacco was observed presumably
because the region around the transcription initiation
site includes the dicot Type I NEP promoter consensus.
As in rice, the Os-PclpP-111 NEP promoter is active in
mature tobacco leaves. Therefore, it belongs to the
relatively small number of Type I NEP promoters which are
active in mature chloroplasts as well as proplastids.

Transcription from the rice clpP 5' region at a
second site, with a short stretch of homologous sequence
to the tobacco PclpP -53 Type II NEP promoter was
unexpected. Since rice contains all the cis elements
required for Type II promoter activity, lack of
transcription in rice should be due to the evolutionary
loss of the specificity factor required for Type II
promoter recognition. This specificity factor is well
conserved during evolution, as evidenced by an active
Type II clpP promoter in the liverwort *Marchantia*
polymorpha and the conifer *Pinus contorta* (Figure 9).

Experiments reported here for the rice
Os-PclpP-111 NEP promoter suggest that the Type I NEP
transcription machinery is sufficiently conserved between
monocots and dicots to ensure faithful recognition of
heterologous promoters. The clpP mRNA accumulates to
significant levels in all plastid types (Shanklin et al.,
1995, *supra*; Hajdukiewicz et al., 1997, *supra*; Example
I). Therefore, the strong, constitutive Os-PclpP-111 NEP
promoter is suitable for the expression of chimeric genes
in a broad range of crops.

EXAMPLE III

**In vivo definition of a Type II promoter, PclpP-53, for
the nuclear encoded plastid RNA polymerase (NEP)**

In tobacco the *clpP* gene is transcribed from three major NEP promoters initiating transcription -511, -173 and -53 nucleotides upstream of the translation initiation codon, and from a PEP promoter (5' end at -95). Transcription from the Type II PclpP-53 NEP promoter is maintained in the green leaves of wild-type tobacco plants. Therefore, given its potential to drive the expression of selectable marker genes, the constitutive PclpP-53 promoter was chosen for analysis.

Materials and Methods for Example III

Construction of Plasmids. Plasmid pPS8 contains a *uidA* reporter gene as a *SacI*-*HindIII* fragment in a pBSKS+ plasmid (Stratagene). The chimeric *uidA* gene consists of: Between the *SacI* and *XhoI* sites, the PclpP-53(-22/+25) promoter fragment containing 22 nt upstream and 25 nt downstream (+1 is nt where transcription initiates) of the *clpP* transcription initiation site; Between *XhoI* and *NcoI* sites, a ribosome binding site; Between the *NcoI* and *XbaI* sites, the *uidA* coding region with an N-terminal c-myc tag corresponding to amino acids 410-419 (EQKLISEEDL; SEQ ID NO: 53) within the carboxy terminal domain of the human c-myc protein (Kolodziej and Young, *Meth. Enz.* **194**:508-519, 1991); Between the *XbaI* and *HindIII* sites the 3' untranslated region of the *rps16* ribosomal protein gene (*Trps16*). DNA sequence of the chimeric *uidA* gene between the *SacI* and *HindIII* sites in plasmid pPS8 is shown in Figure 13. Relevant restriction sites of the chimeric *uidA* gene are shown in Fig. 10, where the *uidA* gene is shown as part of plasmid pPS18. Plasmid pPS18 was obtained by cloning the *uidA* gene as a *SacI*-*HindIII* fragment into *SacI*-*HindIII*-digested pPRV111A plastid transformation vector. Plasmid pPRV111A, a pBSKS+ plasmid derivative (Stratagene), and was described in Zoubenko et al., 1994, *supra*.

Plasmids pPS16, pPS37, pPS17 and pPS38 listed in Figure 11 were obtained from plasmid pPS18 by replacing the PclpP-53(-22/+25) SacI-XhoI promoter fragment with the PclpP-53(-152/+154),
5 PclpP-53(-152/+41), PclpP-53(-152/+10),
PclpP-53(-39/+154) promoters, respectively. The SacI-XhoI fragments were obtained by PCR amplification. PCR primers are listed according to the position of the terminal
10 nucleotide relative to the transcription initiation site (it is the complement of nt 74557 in the tobacco plastid genome: accession no. Z00044):

clpP-152	ccgagctcGAATGAGTCCATACTTAT	SEQ ID NO: 54
clpP-39	ccgagctcAAAACCAATATGAATATTATA	SEQ ID NO: 55
clpP -22	ccgagctcTATAAAGACAATAAAAAAAT	SEQ ID NO: 56
15 clpP+10	ccctcgagAAACGTAACAATTTTTTTT	SEQ ID NO: 57
clpP+25	ccctcgagTTTCACTTTGAGGTGGA	SEQ ID NO: 58
clpP+41	ccctcgagAGAACTAAATACTATATTC	SEQ ID NO: 59
clpP+154	ccctcgagATATGACCCAATATATCTG	SEQ ID NO: 60

20 Anchor sequences derived from the plastid genome are in capital letters ; added nucleotides to create restriction sites (underlined) are in lower case letters.

25 *Tobacco Plastid Transformation.* Plastid transformation and regeneration of tobacco plants was carried out as described by Svab and Maliga, 1993, *supra*. Transgenic plants were selected on regeneration medium containing 500 µg/ml spectinomycin dihydrochloride.

30 *Primer Extension Analysis.* Total leaf RNA was isolated from the leaves of transgenic plants maintained on RM medium, by the method of Stiekema et al., 1988, *supra*. Primer extension reactions were carried out as
35 described by Allison and Maliga, 1995, *supra*, using 15 µg of the total RNA and primer PE1 complimentary to the 5' end of the uidA coding sequence.
Primer PE1 sequence:

5'-GGCCGTCGAGTTTTTTTGATTTCACGGGTGGGG-3' (SEQ ID NO: 61)

Results and Discussion

To identify functionally important sequences in the Type II PclpP-53 promoter, expression of reporter genes driven by sequences surrounding the clpP-53 NEP transcription initiation site was measured. Deletion of sequences from the 5' and 3' ends facilitated the determination of the boundaries of the promoter. These studies revealed that a 47-bp fragment is sufficient to support accurate transcription initiation. The data further suggest that not more than 28 bp out of 47 are essential for promoter function. A majority of the relevant sequences are downstream of the transcription initiation site.

Since all plastid promoters are within about 150 nucleotides of the transcription initiation site, transcription initiation in vivo from a 306-bp fragment (-152/+154) surrounding the clpP -53 transcription initiation site was assessed. For testing promoter function, clpP fragments were cloned upstream of a uidA reporter construct encoding β -glucuronidase. See Figure 11. The uidA construct has a ribosome binding site between the XhoI and NcoI restriction sites, as well as the 3'-untranslated region of the plastid rps16 gene (Trps16) for stabilization of the mRNA. The chimeric uidA gene was introduced into tobacco plastids by linkage to a selectable spectinomycin resistance (aadA) gene. The 306-bp fragment contains a PEP and a NEP promoter (Fig. 11). Functioning of both promoters was established by primer extension analysis. See Figure 12.

After confirming that the 306-bp fragment is sufficient to drive NEP transcription, a series of deletions were made from the 5' as well as 3' end. These constructs were then tested for transcription initiation in vivo in tobacco. The schematic design of the promoter deletions is shown in Fig. 11, the primer extension data are shown in Fig. 12. Primer extension analysis on the

series of deletions showed that sequences extending from -10 to +25 (pPS43; Fig. 12) and -5 to +25 (pPS44; not shown) are sufficient to support accurate initiation from clpP -53. Upon overexposure, a faint band of the proper size is observed even in the +1 to +25 construct (pPS45, not shown). Also, transcription from the NEP promoter was abolished completely in the -152/+10 construct, (pPS17; Fig. 12) and in the -22 to +21 construct, (pPS41; not shown). This indicates that sequences between +5 and +25 are important for transcription from the NEP promoter. Some of this data is summarized in Figs. 11 and 12.

Sequences required for transcription by the NEP polymerase are not known. Based on conservation of the ATAGAATA/GAA around the transcription initiation site, Hajdukiewicz et al., 1997, *supra*, most NEP promoters were classified as Type I. The promoter studied here, PclpP-53, lacks this sequence motif, because of which it is classified as Type II. Transcription analysis in the truncated promoter fragment of plasmid pPS18 in vivo shows that sequences required to support transcription from PclpP-53 are located within a 30 basepair fragment extending from -5 to +25 with respect to the transcription initiation site. Furthermore, nucleotides between +10 and +25 are important for transcription initiation, since there is no transcription from clpP promoter derivatives in plasmid pPS17 and pPS41 lacking this region.

Expression of the rice clpP promoter region revealed a transcript that mapped to a region with homology to the transcription initiation site of the PclpP-53 promoter. Alignment of the rice and the tobacco sequences show significant homology downstream of the transcription initiation site, and not much homology upstream. Considering that this rice sequence was recognized by tobacco in vivo (Example II), sequences that are important for transcription initiation should be

present in the rice sequence. Combining this information with the finding that -22/+25 sequences are sufficient for transcription initiation in vivo, it appears that sequences 5'-ATTGTTACGTTTCCACCTCAAAGTGAAA-3' (portion of SEQ ID NO: 25) extending from -3 to +25 contain the information which is important for PclpP-53 promoter function.

Since the constitutive Type II PclpP-53 is efficiently transcribed in all tissue types, it is useful for the expression of selectable marker genes, and of proteins of economic value in all dicot plants.

EXAMPLE IV

PLASTID PROMOTER UTILIZATION IN RICE EMBRYOGENIC CELL CULTURE

The utilization of the clpP promoter in tobacco and maize plastids is described in the previous examples. The present example is directed to the analysis of plastid promoter utilization in rice. The 5' ends of several mRNA species were mapped in samples derived from cultured embryogenic rice cells and leaves. The RNAs for clpP and 16SrRNA are relatively abundant in embryogenic rice cells indicating that the promoters of these genes may be used to advantage to drive plastid expression of selectable marker genes and/or foreign genes of interest in rice.

Plastid transformation in rice is highly desirable. The present example provides compositions and methods to effectuate rice plastid transformation in embryogenic cultured rice cells. Such cells may be efficiently regenerated into mature plants (Vasil, IK (1994) Plant Mol Biol 25:925-937; Christou, P (1996) Trends Plant Sci 1:423-431). Data from cultured tobacco (BY2) cells suggest that plastid promoter utilization in tissue culture may be different from those in leaves (Vera A, Sugiura M (1995) Curr Genet 27:280-284; Vera A, Hirose T, Sugiura M (1996) Mol Gen Genet 251:518-525

et al., 1996; Kapoor S, Suzuki JY, Sugiura M (1997) Plant J 11:327-337). The data presented herein reveal that cultured embryogenic rice cells and leaves utilize the same promoter. *rbcL*, *atpB*, 16SrDNA and *clpP* have only one promoter each which is recognized by the plastid-encoded plastid RNA polymerase (PEP). In contrast, *clpP* is transcribed by the nucleus-encoded plastid RNA polymerase (NEP) in both samples. The RNAs for *clpP* and 16SrRNA are relatively abundant in embryogenic cells indicating that the promoters of these genes may be suitable to drive the expression of selectable marker genes.

Materials and methods for Example IV

Plant Materials. Embryogenic rice callus was initiated from mature seed of cv. Taipei 309 on LS2.5 medium (Abdullah R, Cocking EC, Thompson JA (1986) Bio/Technology 4:1088-1090). The calli were introduced into liquid AA medium (Muller AJ, Grafe R (1978) Mol Gen Genet 161: 67-76) to establish embryogenic suspension cultures, and subcultured at biweekly intervals. DNA and RNA were prepared from 3 month old cultures 14 days after subculture. Plants were regenerated from embryogenic calli on complete MS medium supplemented with 2 mg/L BAP and 3% sucrose (Murashige T, Skoog F (1962) Physiol Plant 15: 473-497 1962) and transferred onto hormone-free MS medium. Leaves for the isolation of nucleic acids were taken from these plants after four months.

RNA and DNA gel blots. Total cellular RNA was prepared according to Stiekema et al., 1988, *supra*. The RNA (5 µg per lane) was subjected to electrophoresis in a formaldehyde-agarose gel, blotted and hybridized (Hajdukiewicz PTJ, Allison LA, Maliga P (1997) EMBO J 16:4041-4048). Double-stranded ptDNA probes were prepared by random-primed ³²P-labeling of PCR-generated or gel-purified DNA fragments. The sequence of the primers

used for PCR, along with their positions within the tobacco (N.t.; accession no. Z00044; Shinozaki et al. 1986, *supra*) or maize (Z.m.; accession no. X86563; Maier et al. 1995, *supra*) ptDNA are as follows:

Gene	5' nt position in plastid DNA	Sequence	SEQ ID NO:
atpB (Z.m.)	55860 (C)	GAGAGGAATGGAAGTGATTGACA	(62)
	55103	GAGCAGGGTCGGTCAAATC	(63)
clpP (Z.m.)	69840	ATCCTAGCGTGAGGGAATGCTA	(64)
	70064 (C)	AGGTCTGATGGTATATCTCAGTAT	(65)

The following ptDNA fragments were used as probes: *rbcL* (N.t.), a BamHI fragment (nucleotides 58047 to 59285 in ptDNA); 16SrDNA (N.t.), EcoRI to EcoRV fragment (nucleotides 138447 to 140855).

The probe for tobacco 25S rRNA was from plasmid pKDR1 (Dempsey et al., 1993, *supra*) containing a 3.75 kb EcoRI fragment from a tobacco 25S/18S locus cloned in plasmid pBR325.

Total leaf DNA for relative plastid genome copy number determination was prepared (Mettler IJ (1987) Plant Mol Biol Rep 5:346-349), digested with the EcoRI restriction endonuclease, separated on 0.7% agarose gels, blotted and hybridized with the plastid 16SrDNA and cytoplasmic 25SrDNA probes (Allison et al., 1996, *supra*).

Primer extension analysis. Primer extension reactions were carried out on total leaf RNA as described (Allison and Maliga, 1995, *supra*). The primers are listed below, with nucleotide position in the published rice plastid genome sequence (Hiratsuka et al., 1989, *supra*).

Gene	5' nt position in plastid DNA	Sequence	SEQ ID NO:
<i>rbcL</i>	54124 (C)	ACTTGCTTTAGTTTCTGTTTGTGGTGACAT	(66)
<i>atpB</i>	53287	AGAAGTAGTAGGATTGGTTCTCATAAT	(67)
16S rRNA	123777	CCGCCAGCGTTCATCCTGAGC	(68)
<i>clpP</i>	68263	GGTACTTTTGGAACACCAATGGGCAT	(69)

Primer extension reactions were carried out with 1 µg of RNA from leaves, and 10 µg (clpP, 16SrDNA) and 30 µg (rbcL, atpB) of RNA from embryogenic cells.

Results and Discussion

Several plastid promoters have been identified in rice and related monocots. To assess whether these promoters would be suitable to drive expression of selectable marker genes and/or foreign genes of interest in these plant species, transcript accumulation was examined. The rbcL gene in rice and maize is transcribed from a PEP promoter (Mullet et al., 1985, *supra*; Nishizawa Y, Hirai A (1987) Jpn J Genet 62:389-395). The atpB gene in maize chloroplasts is transcribed from a PEP promoter (Mullet et al., 1985, *supra*), whereas in maternal white iojap seedlings lacking PEP it is transcribed from an alternative NEP promoter. The 16SrDNA gene (the first gene of the plastid ribosomal RNA operon) in barley chloroplasts is transcribed from a PEP promoter (Reinbothe S, Reiboth C, Heintzen C, Seidenbecher C, Parthier B (1993) EMBO J 12:1505-1512), whereas in the white albobistrans seedlings lacking PEP it is transcribed from an uncharacterized NEP promoter (Hess et al., 1993, *supra*). The clpP gene in wild-type and iojap maize chloroplasts is transcribed from a NEP promoter.

To determine the level of expression of these genes in embryogenic rice cells and leaf cells, accumulation of mRNAs was assessed on Northern blots. The data reveal that transcript levels in embryogenic cells relative to leaves were barely detectable for rbcL (153-times lower), reduced for atpB (37-fold lower) and 16SrDNA (7-fold lower), and similar for clpP (approximately 1.1-times lower). See Figure 14. Interestingly, the number of plastid genome copies (ptDNA) per cell is about the same in embryogenic cells and leaves. See Figure 15. Accordingly, the differences

in transcript levels represent values normalized for
ptDNA copy number.

To further characterize active plastid
promoters, transcript 5' ends were mapped by primer
extension in cultured embryogenic cells and in leaves.
Two different 5' ends, at 312 and 58 nucleotides upstream
of the translation initiation codon, were mapped for
rbcL. See Figure 16. The same two 5'-ends were
identified in leaf chloroplasts, and in the plastids of
embryogenic cells. Two 5'-ends were mapped to similar
positions by S1 nuclease analysis in rice chloroplasts
(Nishizawa and Hirai, 1987, *supra*). The rbcL -312-end
is downstream of -10/-35 σ^{70} -type promoter elements which
are conserved in monocots; the -58-end is generated by
RNA processing (Mullet et al., 1985, *supra*; Reinbothe et
al., 1993, *supra*).

For atpB, a single 5'-end was mapped 310
nucleotides upstream of the translation initiation codon.
As for rbcL, the same 5'-ends were identified in
embryogenic cells and leaves. See Figure 16. The -310
end is associated with PEP promoter elements, and has
been reported earlier in chloroplasts for rice (Nishizawa
Y, Hirai A (1989) *Jpn J Genet* 64:223-229) and maize
(Mullet et al., 1985, *supra*).

For the rRNA operon, the same two 5' ends were
mapped upstream of the mature 16SrRNA in embryogenic
cells and in leaves shown in Figure 16. Based on DNA
sequence conservation, the -116 end is the product of a
PEP promoter whereas the -28 end derives from RNA
processing. See Figure 17A; Strittmatter G,
Godzicka-Josefiak A, Kossel H (1985) *EMBO J* 4:599-604;
Vera and Sugiura, 1995, *supra*; Allison et al., 1996,
supra).

For clpP, the same 5' end was mapped in
embryogenic tissue culture cells and in leaves. See
Figure 16. The transcript initiates 111 nucleotides
upstream of the translation initiation codon within the

10-nucleotide NEP consensus, shown in Figure 5A, and is the product of the clpP Type-I NEP promoter.

Mapping of RNA 5' ends upstream of rbcL, atpB, 16SrDNA and clpP identified the same promoters in cultured embryogenic cells and in the leaves of rice. The data obtained from these studies in rice can be contrasted with the results reported for tobacco. In tobacco, atpB and 16SrDNA are preferentially transcribed from PEP promoters in leaves, and from NEP promoters in BY2 tissue culture cells (Vera and Sugiura 1995; Kapoor et al. 1997). In rice embryogenic cultures, no PEP to NEP promoter switch was observed for any of the plastid genes examined. The data suggest that a PEP to NEP promoter switch is not essential for adaptation to cell culture. Alternatively, atpB and 16SrDNA, genes which have PEP and NEP promoters in other monocots as discussed in the previous examples, have no NEP promoters in rice.

One important difference between embryogenic rice cells and the tobacco BY2 cell line is the length of time in culture. The rice cell culture line described herein is only a few month old and has maintained the ability to regenerate plants. The BY2 cell line has been grown in culture for several years and has lost the capacity for plant regeneration (Yasuda T, Kuroiwa T, Nagata T (1988) Planta 174:235-241). Accordingly the possibility remains that in BY2 cells plastid gene expression may be different from tobacco cells in a short-term culture or in a tobacco plant.

The results presented herein provide practical implications for the production of transgenic rice cells and or plants. Two powerful promoters are described herein which are active in rice embryogenic cells: the clpP NEP promoter and the rrn PEP promoter. Both promoters are suitable to drive the expression of selectable marker genes for plastid transformation in embryogenic rice cells.

While certain of the preferred embodiments of

the present invention have been described and specifically exemplified above, it is not intended that the invention be limited to such embodiments. Various modifications may be made thereto without departing from the scope and spirit of the present invention, as set forth in the following claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Maliga, Pal
Sriraman, Priya
Silhavy, Daniel

(ii) TITLE OF INVENTION: Plastid Promoters for Transgene Expression in the Plastids of Higher Plants

(iii) NUMBER OF SEQUENCES: 69

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Dann, Dorfman, Herrell and Skillman
(B) STREET: 1601 Market Street Suite 720
(C) CITY: Philadelphia
(D) STATE: PA
(E) COUNTRY: USA
(F) ZIP: 19103-2307

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: Not yet assigned
(B) FILING DATE: 03-JUN-1998
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: 60/048,376
(B) FILING DATE: 03-JUN-1997

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: 60/058,670
(B) FILING DATE: 12-SEP-1997

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Rigaut, Kathleen D.
(B) REGISTRATION NUMBER: P43,047

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (215) 563-4100
(B) TELEFAX: (215) 563-4044

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 149 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GACTGTTTTA TCAATTCATT TTTATTCCAT TTCAACCCCT GCTAAATTCG AACTTTCGTC 60
GAAATCGTCT CTATTCATAT GTATGAAATA CATATATGAA ATACGTATGT GGAGTTCCT 120
AGAATTCAT GTGATTCAGT AACAGAAT 149

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 149 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: not relevant

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

TTGCAAAAT CTAAAAA TGAATTTAA TTAATATCAA CTCATTAAAT AAAAAAGGA 60
GTATGCTTAA GTTAATGAAT ATGTTTCATT CATATATAAT GTGTACACCC TGTGTACGTT 120
CTATCCTATA GGAATTTTAC TATAGGAAT 149

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 149 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: not relevant

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ATCACGGATT CTTTTTCTT TATTCAATCT GTTTTACCTT CCTTATATGT AGAATATTTC 60
AATCTATGTA TTAATAGAAT CTATAGTATT CTTATAGAAT AAGAAAAA AAATGAAGAT 120
AATAAAGTGC GGATTCTTTC TTTCTCTTC 149

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 41 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: DNA (genomic)

- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

5 TAAGTTAATG AATATGTTTC ATTCATATAT AATGTGACAC C 41

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 41 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: not relevant
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (iii) HYPOTHETICAL: NO
 - (iv) ANTI-SENSE: NO
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:
- 10 TAAGTTAATG AATATGTTTC ATTCATATAT AATGTGACAC C 41

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 41 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: not relevant
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (iii) HYPOTHETICAL: NO
 - (iv) ANTI-SENSE: NO
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:
- 20 TAGGTTAATG AATATGTTTC ATTCATATAT AATGCGACAC C 41

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 41 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: not relevant
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (iii) HYPOTHETICAL: NO
 - (iv) ANTI-SENSE: NO
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:
- 30 TAGGTTAATG AATATGTTTC ATTCATATAT AATGCGACAC C 41

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 41 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: not relevant
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (iii) HYPOTHETICAL: NO
 - (iv) ANTI-SENSE: NO
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:
- 45 TCATTCATAT AATATGTTTC ATTCATATAT AATGGGACAC C 41

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 41 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: not relevant
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (iii) HYPOTHETICAL: NO
 - (iv) ANTI-SENSE: NO
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:
- 60
- 65

CTCTATTTCAT ATGTATGAAA TACATATATG AAATACGTAT G 41

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 41 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

CTCTATTTCAT ATGTATGAAA TACATATATG AAATACGTAT G 41

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 39 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

CAGGTTGGAA TGTGTATTAT CATAATAATG GTAGAAATG 39

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 41 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

TTAATAGAAT CTATAGTATT CTTATAGAAT AAGAAAAAAAA A 41

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 41 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

TTAATAGAAT CTATAGTATT CATATAGAAT AAGAAAAAAAA C 41

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 41 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

TTAATAGAAT CTATAGTATT CATATAGAAT AAGAATAAAA T 41

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 251 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

```
TCGAATCACC ATTCTTTTTT CTTTATTCAA TCTGCTCTAT CCTACTTATA TGTATAATCT 60
TTCAATCTAT GTATTATTTT AATCTACGTA CTTAATAGAA TCTATAGTAT TCATATAGAA 120
TAAGAAAAAA ACGTGAAAAC AATAAACTGC GGATTCTTTC TTTCTCTTCC ATTCTTACGT 180
TTCCATATTA AAGTGTAGTT TTCTTACTTA AATTTAATAA TATTAATCTA ATATGCCCAT 240
TGGTGTTCCA A                                     251
```

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 199 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

```
TAGAAAGACC TATTCGTAAT AATTTGAGTT TATTCATTCT GTCTTTCTTT ATGAATTTTT 60
ATAATCTATG GATAAAATAA ATACGATAAA AACCAATATG AATATTATAA AGACAATAAA 120
AAAAATTGTT ACGTTTCCAC CTCAAAGTGA AATATAGTAT TTAGTTCCTTT CTTTCATTTA 180
ATGCCTATTG GTGTTCCAA                                     199
```

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 283 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

```
GAGCTCGAAT CACCATTCTT TTTTCTTTAT TCAATCTGTC TTATCCTACT TATATGTATA 60
ATCTTTCAAT CTATGTATTA TTTCAATCTA CGTACTTAAT AGAATCTATA GTATTCATAT 120
AGAATAAGAA AAAAACGTGA AAACAATAAA CTGCGGATTC TTTCTTTCTC TTCCATTCTT 180
ACGTTTCCAT ATTAAAGTGT AGTTTTCTTA CTAAATTTTA ATAATATTAA TCTAATATGC 240
CCATTGGTGT TCCAAGAATT CAGTTGTAGG GAGGGATCCA TGG                                     283
```

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 77 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

```
TAAATAAATA GAATTTTCATT TTTACGTTTT TTTATTATAG AAGAGTATTT TGTTTGTGGA 60
AGAAAAAATA AATGCCT                                     77
```

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 78 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: not relevant
5 (ii) MOLECULE TYPE: DNA (genomic)
(iii) HYPOTHETICAL: NO
(iv) ANTI-SENSE: NO
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

10 TGTTACACAA CTTCATATAC TTTACGTTCC CATATTATAG TATAGTGCT TAACTTCTTT 60
CCATTAAAAC AAATGCCC 78

(2) INFORMATION FOR SEQ ID NO:20:
15 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 80 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: not relevant
20 (ii) MOLECULE TYPE: DNA (genomic)
(iii) HYPOTHETICAL: NO
(iv) ANTI-SENSE: NO
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

25 TAAAGACAAT AACCGTAATT ATTACGTTTC CACATCAAAG TGAAATAGAG TACTTAATTT 60
TTTTCTTTCA TTTAATGCCT 80

(2) INFORMATION FOR SEQ ID NO:21:
30 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 79 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: not relevant
35 (ii) MOLECULE TYPE: DNA (genomic)
(iii) HYPOTHETICAL: NO
(iv) ANTI-SENSE: NO
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

40 TAAAGACAAT AAAAAAATTT GTTACGTTTC CACCTCAAAG TGAAATATAG TATTTAGTTC 60
TTTCTTTTCAT TTAATGCCT 79

(2) INFORMATION FOR SEQ ID NO:22:
45 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 82 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: not relevant
50 (ii) MOLECULE TYPE: DNA (genomic)
(iii) HYPOTHETICAL: NO
(iv) ANTI-SENSE: NO
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

55 TTCTTTCTTT CTCTTCCATT CTTACGTTTC CATATTAAAG TGTAGTTTTT TACTTAAAT 60
TTAATAATAT TAATCTAATA TG 82

(2) INFORMATION FOR SEQ ID NO:23:
60 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 82 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: not relevant
65 (ii) MOLECULE TYPE: DNA (genomic)
(iii) HYPOTHETICAL: NO
(iv) ANTI-SENSE: NO
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

TTCTTTCTTT CTCTTCCATT CTTACGTTTC CATATTAAAG TGTAGTTTTT TACTTAAAT 60

TTAATAATAT TAATCTAATA TG

82

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 83 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

TTAAAAAACG AAACCCCAAT TTTACGTTTC CACATCAAAG TGAAATAGAG AACTTCATTC 60
TCTTTTTTTTT TCATTTTCATG CCT 83

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 59 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

GAGCTCTATA AAGACAATAA AAAAAATTGT TACGTTTCCA CCTCAAAGTG AAAC TCGAG 59

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

AAAAAAAATT GTTACGTTTC CACCTCAAAG TGAAA 35

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2141 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GAGCTCTATA	AAGACAATAA	AAAAAATTGT	TACGTTTCCA	CCTCAAAGTG	50
AAACTCGAGA	ATTCAGTTGT	AGGGAGGGAT	CCATGGAACA	AAAACTCATT	100
TCTGAAGAAG	ACTTGGTACG	TCCTGTAGAA	ACCCCAACCC	GTGAAATCAA	150
AAACTCGAC	GGCCTGTGGG	CATTCACTCT	GGATCGCGAA	AACTGTGGAA	200
TTGATCAGCG	TTGGTGGGAA	AGCGCGTTAC	AAGAAAGCCG	GGCAATTGCT	250
GTGCCAGGCA	GTTTTAACGA	TCAGTTCGCC	GATGCAGATA	TTTCGTAATTA	300
TGCGGGCAAC	GTCTGGTATC	AGCGCGAAGT	CTTTATACCG	AAAGGTGTTGG	350
CAGGCCAGCG	TATCGTGCTG	CGTTTCGATG	CGGTCACCTA	TTACGGCAAA	400
GTGTGGGTCA	ATAATCAGGA	AGTGATGGAG	CATCAGGGCG	GCTATACGCC	450
ATTTGAAGCC	GATGTCACGC	CGTATGTTAT	TGCCGGGAAA	AGTGATCGTA	500
TCACCGTTTG	TGTGAACAAC	GAAGTGAAGT	GGCAGACTAT	CCCGCCGGGA	550
ATGGTGATTA	CCGACGAAAA	CGGCAAGAAA	AAGCAGTCTT	ACTTCCATGA	600
TTTCTTTAAC	TATGCCGGAA	TCCATCGCAG	CGTAATGCTC	TACACCACGC	650
CGAACACCTG	GGTGGACGAT	ATCACCGTGG	TGACGCATGT	CGCGCAAGAC	700

```

5      TGTAACCACG CGTCTGTTGA CTGGCAGGTG GTGGCCAATG GTGATGTCAG 750
      CGTTGAACTG CGTGATGCGG ATCAACAGGT GGTTGCAACT GGACAAGGCA 800
      CTAGCGGGAC TTTGCAAGTG GTGAATCCGC ACCTCTGGCA ACCGGGTGAA 850
      GGTATCTCT ATGAACGTGT CGTCACAGCC AAAAGCCAGA CAGAGTGTGA 900
      TATCTACCCG CTTGCGGTCG GCATCCGGTC AGTGGCAGTG AAGGGCCAAC 950
      AGTTCCTGAT TAACCACAAA CCGTTCCTACT TTACTGGCTT TGGTCGTCAT 1000
      GAAGATGCGG ACTTACGTGG CAAAGGATTC GATAACGTGC TGATGGTGCA 1050
      CGACCACGCA TTAATGGACT GGATTGGGGC CAACTCCTAC CGTACCTCGC 1100
      ATTACCCTTA CGCTGAAGAG ATGCTCGACT GGGCAGATGA ACATGGCATC 1150
10     GTGGTGATTG ATGAAACTGC TGCTGTCGGC TTTAACCCTCT CTTTAGGCAT 1200
      TGGTTTCGAA GCGGGCAACA AGCCGAAAGA ACTGTACAGC GAAGAGGCAG 1250
      TCAACGGGGA AACTCAGCAA GCGCACTTAC AGGCGATTAA AGAGCTGATA 1300
      GCGCGTGACA AAAACCACCC AAGCGTGGTG ATGTGGAGTA TTGCCAACGA 1350
      ACCGGATACC CGTCCGCAAG TGCACGGGAA TATTTGCGCA CTGGCGGAAG 1400
15     CAACGCGTAA ACTCGACCCG ACGCGTCCGA TCACCTGCGT CAATGTAATG 1450
      TTCTGCGACG CTCACACCGA TACCATCAGC GATCTCTTTG ATGTGCTGTG 1500
      CCTGAACCGT TATTACGGAT GGTATGTCCA AAGCGGCGAT TTGGAAACGG 1550
      CAGAGAAGGT ACTGGAAAAA GAACTTCTGG CCTGGCAGGA GAAACTGCAT 1600
      CAGCCGATTA TCATCACCGA ATACGGCGTG GATACGTTAG CCGGGCTGCA 1650
20     CTCAATGTAC ACCGACATGT GGAGTGAAGA GTATCAGTGT GCATGGCTGG 1700
      ATATGTATCA CCGCGTCTTT GATCGCGTCA GCGCCGTCGT CCGTGAACAG 1750
      GTATGGAATT TCGCCGATTT TGCGACCTCG CAAGGCATAT TCGCGCTTGG 1800
      CCGTAACAAG AAAGGGATCT TCACTCGCGA CCGCAAACCG AAGTCGGCGG 1850
      CTTTCTGCT GCAAAAACGC TGGACTGGCA TGAACCTCGG TGAAAAACCG 1900
25     CAGCAGGGAG GCAAAACAATG AATCAACAAC TCTCTGGCG CACCATCGTC 1950
      GGCTACAGCC TCGGTGGGGA ATTGCTCTAG AGAAATTCAA TTAAGGAAAT 2000
      AAATTAAGGA AATACAAAAA GGGGGGTAGT CATTTGTATA TAACTTTGTA 2050
      TGACTTTTCT CTTCTATTTT TTTGTATTTT CTCCCTTTCC TTTTCTATTT 2100
30     GTATTTTTTT ATCATTGCTT CCATTGAATT AATTCAAGCT T 2141

```

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 200 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

```

CACCACGATC GAACGGGAAT GGATAGGAGG CTTGTGGGAT TGACGTGATA GGGTAGGGTT 60
GGCTATACTG CTGGTGGCGA ACTCCAGGCT AATAATCTGA AGCGCATGGA TACAAGTTAT 120
CCTTGGAAGG AAAGACAATT CCGAATCCGC TTTGTCTACG AATAAGGAAG CTATAAGTAA 180
TGCAACTATG AATCTCATGG                                     200

```

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 200 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

```

CGCCACGATC GAACGGGAAT GGATAAGAGG CTTGTGGGAT TGACGTGATA GGGTAGGGTT 60
GGCTATACTG CTGGTGGCGA ACTCCAGGCT AATAATCTGA AGCGCATGGA TACAAGTTAT 120
CCTTGGAAGG AAAGACAATT CCGAATCCGC TTTGTCTACG AATAAGGAAG CTATAAGTAA 180
TGCAACTATG AATCTCATGG                                     200

```

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 61 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: not relevant
5 (ii) MOLECULE TYPE: DNA (genomic)
(iii) HYPOTHETICAL: NO
(iv) ANTI-SENSE: NO
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

TTAATAGAAT CTATAGTATT CTTATAGAAT AAGAAAAAAA AAATGAAGAT AATAAACTGC 60
G 61

(2) INFORMATION FOR SEQ ID NO:31:
15 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 60 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: not relevant
(ii) MOLECULE TYPE: DNA (genomic)
20 (iii) HYPOTHETICAL: NO
(iv) ANTI-SENSE: NO
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

TTAATAGAAT CTATAGTATT CATATAGAAT AAGAAAAAAA CGTGAAAACA ATAACTGCG 60

25 (2) INFORMATION FOR SEQ ID NO:32:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 133 base pairs
(B) TYPE: nucleic acid
30 (C) STRANDEDNESS: single
(D) TOPOLOGY: not relevant
(ii) MOLECULE TYPE: DNA (genomic)
(iii) HYPOTHETICAL: NO
(iv) ANTI-SENSE: NO
35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

GCTCCCCCGC CGTCGTTCAA TGAGAATGGA TAAGAGGCTC GTGGGATTGA CGTGAGGGGG 60
CAGGGATGGC TATATTCTGG GAGCGAACTC CGGGCGAATA CGAAGCGCTT GGATACAGTT 120
GTAGGGAGGG ATT 133

40 (2) INFORMATION FOR SEQ ID NO:33:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
45 (C) STRANDEDNESS: single
(D) TOPOLOGY: not relevant
(ii) MOLECULE TYPE: DNA (genomic)
(iii) HYPOTHETICAL: NO
(iv) ANTI-SENSE: NO
50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

GAGAGGAATG GAAGTGATTG ACA 23

(2) INFORMATION FOR SEQ ID NO:34:
55 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: not relevant
60 (ii) MOLECULE TYPE: DNA (genomic)
(iii) HYPOTHETICAL: NO
(iv) ANTI-SENSE: NO
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

GAGCAGGGTC GGTCAAATC 19

65 (2) INFORMATION FOR SEQ ID NO:35:
(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: not relevant
5 (ii) MOLECULE TYPE: DNA (genomic)
(iii) HYPOTHETICAL: NO
(iv) ANTI-SENSE: NO
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

10 ATCCTAGCGT GAGGGAATGC TA 22

(2) INFORMATION FOR SEQ ID NO:36:
(i) SEQUENCE CHARACTERISTICS:
15 (A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: not relevant
(ii) MOLECULE TYPE: DNA (genomic)
20 (iii) HYPOTHETICAL: NO
(iv) ANTI-SENSE: NO
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

AGGTCTGATG GTATATCTCA GTAT 24

25 (2) INFORMATION FOR SEQ ID NO:37:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
30 (C) STRANDEDNESS: single
(D) TOPOLOGY: not relevant
(ii) MOLECULE TYPE: DNA (genomic)
(iii) HYPOTHETICAL: NO
(iv) ANTI-SENSE: NO
35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

CGCTTCTGTA ACTGG 15

(2) INFORMATION FOR SEQ ID NO:38:
(i) SEQUENCE CHARACTERISTICS:
40 (A) LENGTH: 16 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: not relevant
(ii) MOLECULE TYPE: DNA (genomic)
45 (iii) HYPOTHETICAL: NO
(iv) ANTI-SENSE: NO
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

50 TGACTGTCAA CTACAG 16

(2) INFORMATION FOR SEQ ID NO:39:
(i) SEQUENCE CHARACTERISTICS:
55 (A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: not relevant
(ii) MOLECULE TYPE: DNA (genomic)
(iii) HYPOTHETICAL: NO
(iv) ANTI-SENSE: NO
60 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

GGTACTTTTG GAACACCAAT GGGCAT 26

65 (2) INFORMATION FOR SEQ ID NO:40:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

GAAGTAGTAG GATTGGTTCT CATAAT 26

- (2) INFORMATION FOR SEQ ID NO:41:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: not relevant
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (iii) HYPOTHETICAL: NO
 - (iv) ANTI-SENSE: NO
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

GGTCTAGAAT TCCTATCGAA TTCCTTC 27

- (2) INFORMATION FOR SEQ ID NO:42:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: not relevant
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (iii) HYPOTHETICAL: NO
 - (iv) ANTI-SENSE: NO
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

GAATCTACAA AATCCCTCGA ATTG 24

- (2) INFORMATION FOR SEQ ID NO:43:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: not relevant
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (iii) HYPOTHETICAL: NO
 - (iv) ANTI-SENSE: NO
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

ACTCTTCATC AATCCCTACG 20

- (2) INFORMATION FOR SEQ ID NO:44:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 25 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: not relevant
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (iii) HYPOTHETICAL: NO
 - (iv) ANTI-SENSE: NO
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

GGTCTAGACT AACTTTAAT ATGGA 25

- (2) INFORMATION FOR SEQ ID NO:45:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: not relevant

- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

GGGAATTCTG TTTGTAAGAA GA 22

(2) INFORMATION FOR SEQ ID NO:46:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

GGTCTAGAAT TCCTATCGAA TTCCTTC 27

(2) INFORMATION FOR SEQ ID NO:47:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 29 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

GGCTCGAGGG ACAACTCGAT AGGATTAGG 29

(2) INFORMATION FOR SEQ ID NO:48:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

GGTCTAGAAT CTAGCAATCA TGGAATC 27

(2) INFORMATION FOR SEQ ID NO:49:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 26 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

GGCTCGAGCG TGCTATTCTA AATCGT 26

(2) INFORMATION FOR SEQ ID NO:50:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 23 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO

- (iv) ANTI-SENSE: NO
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

GGGAGCTCGA ATCACCATTC TTT 23

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: not relevant

- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO

- (iv) ANTI-SENSE: NO
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

GGGAATTCTT GGAACACCAA TGGGCAT 27

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: not relevant

- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO

- (iv) ANTI-SENSE: NO
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

GGCCGTCGAG TTTTGTGATT TCACGGGTTG GGG 33

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 amino acids
- (B) TYPE: amino acids
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: not relevant

- (ii) MOLECULE TYPE: peptide
- (iii) HYPOTHETICAL: NO

- (iv) ANTI-SENSE: NO
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

Glu Gln Lys Leu Ile Ser Glu Glu Asp Leu
1 5 10

(2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: not relevant

- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO

- (iv) ANTI-SENSE: NO
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

CCGAGCTCGA ATGAGTCCAT ACTTAT 26

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: not relevant

- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO

- (iv) ANTI-SENSE: NO
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

CCGAGCTCAA AACCAATATG AATATTATA 29

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 29 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

CCGAGCTCTA TAAAGACAAT AAAAAAAT 29

(2) INFORMATION FOR SEQ ID NO:57:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

CCCTCGAGAA ACGTAACAAT TTTTTTT 27

(2) INFORMATION FOR SEQ ID NO:58:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

CCCTCGAGTT TCACTTTGAG GTGGA 25

(2) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 28 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

CCCTCGAGAG AACTAAATAC TATATTTC 28

(2) INFORMATION FOR SEQ ID NO:60:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

CCCTCGAGAT ATGACCCAAT ATATCTG 27

(2) INFORMATION FOR SEQ ID NO:61:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 33 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: not relevant
 (ii) MOLECULE TYPE: DNA (genomic)
 (iii) HYPOTHETICAL: NO
 (iv) ANTI-SENSE: NO
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

GGCCGTCGAG TTTTTTGATT TCACGGGTTG GGG 33

(2) INFORMATION FOR SEQ ID NO:62:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 23 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: not relevant
 (ii) MOLECULE TYPE: DNA (genomic)
 (iii) HYPOTHETICAL: NO
 (iv) ANTI-SENSE: NO
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

GAGAGGAATG GAAGTGATTG ACA 23

(2) INFORMATION FOR SEQ ID NO:63:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 19 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: not relevant
 (ii) MOLECULE TYPE: DNA (genomic)
 (iii) HYPOTHETICAL: NO
 (iv) ANTI-SENSE: NO
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

GAGCAGGGTC GGTCAAATC 19

(2) INFORMATION FOR SEQ ID NO:64:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 22 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: not relevant
 (ii) MOLECULE TYPE: DNA (genomic)
 (iii) HYPOTHETICAL: NO
 (iv) ANTI-SENSE: NO
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

ATCCTAGCGT GAGGGAATGC TA 22

(2) INFORMATION FOR SEQ ID NO:65:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 24 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: not relevant
 (ii) MOLECULE TYPE: DNA (genomic)
 (iii) HYPOTHETICAL: NO
 (iv) ANTI-SENSE: NO
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

AGGTCTGATG GTATATCTCA GTAT 24

(2) INFORMATION FOR SEQ ID NO:66:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

ACTTGCTTTA GTTTCTGTTT GTGGTGACAT 30

- (2) INFORMATION FOR SEQ ID NO:67:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: not relevant
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (iii) HYPOTHETICAL: NO
 - (iv) ANTI-SENSE: NO
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

AGAAGTAGTA GGATTGGTTC TCATAAT 27

- (2) INFORMATION FOR SEQ ID NO:68:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: not relevant
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (iii) HYPOTHETICAL: NO
 - (iv) ANTI-SENSE: NO
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

CCGCCAGCGT TCATCCTGAG C 21

- (2) INFORMATION FOR SEQ ID NO:69:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 26 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: not relevant
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (iii) HYPOTHETICAL: NO
 - (iv) ANTI-SENSE: YES
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

GGTACTTTTG GAACACCAAT GGGCAT 26